

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Kelahiran bayi prematur di Indonesia masih menjadi salah satu kasus yang sangat tinggi. Data dari beberapa rumah sakit menunjukkan presentasi antara 14-20% dari seluruh bayi yang dirawat. Keadaan ini terutama disebabkan masalah sosial-ekonomi yang dialami sebagian besar masyarakat Indonesia [1]. Begitu juga dengan di dunia, menurut data Organisasi Kesehatan Dunia (WHO) menyatakan, 44 persen kematian bayi di dunia pada 2012 terjadi pada 28 hari pertama kehidupan (masa neonatal). Penyebab terbesar (37 persen) ialah kelahiran prematur. Prematur menjadi penyebab kematian kedua tersering pada balita setelah pneumonia.

Bayi prematur masih menjadi masalah besar di Indonesia. Faktanya, berdasarkan laporan Perserikatan Bangsa-Bangsa (PBB) berjudul *Born too Soon, The Global Action Report on Preterm Birth* menyebutkan, secara global 15 juta bayi lahir prematur setiap tahun. Bahkan, lebih dari satu juta bayi meninggal karena komplikasi akibat lahir prematur. Dokter Spesialis Anak Konsultan Bayi Prematur, Dr Rinawati Rohsiswatmo SpA (K), menyebutkan, Indonesia menempati peringkat kelima negara dengan kasus kelahiran bayi prematur terbanyak di dunia. Hal tersebut dapat dilihat berdasarkan jumlah penduduk Indonesia sekitar 255 juta jiwa, dengan presentase angka kelahiran bayi prematur yang mencapai sebanyak 675.700 kasus per tahunnya dari sekitar 4,5 juta kelahiran bayi per tahun [2].

Persalinan prematur dapat terjadi ketika kehamilan belum memasuki minggu ke 37, atau tiga minggu atau lebih sebelum hari perkiraan lahir (hpl). Bayi yang dilahirkan secara prematur biasanya memiliki berat badan kurang dari 2,5 kilogram hal ini yang menyebabkan organ-organ tubuh bayi tidak berfungsi dengan baik dapat menderita berbagai masalah kesehatan yang lebih serius dibandingkan bayi-bayi yang dilahirkan sesuai jadwal kelahirannya. Hingga saat ini, ada beberapa kasus persalinan prematur yang belum diketahui penyebabnya. Ada beberapa faktor dan masalah kesehatan yang dapat memicu persalinan prematur yaitu ibu yang tidak sehat, merokok, riwayat kehamilan, kondisi janin, kondisi psikologis [3].

Dari data yang di perlihatkan diatas oleh Organisasi Kesehatan Dunia, data dari beberapa rumah sakit dan laporan Perserikatan Bangsa – Bangsa adalah suatu bukti

kelahiran bayi prematur merupakan salah satu penyakit yang berbahaya bagi korbannya untuk kelangsungan hidupnya dan bayi yang lahir secara prematur diprediksi tetap terjadi disetiap tahunnya. Pencegahan dan pengendalian terhadap kelahiran premature juga telah banyak dilakukan oleh badan kesehatan dan masyarakat diIndonesia. Akan tetapi masih saja masyarakat yang melahirkan bayi premature setiap tahunnya karena berbagai masalah yang dialami oleh masyarakat terutama masalah ekonomi yang disampaikan seperti diatas karena masih mahal nya biaya dalam dunia kesehatan dan sering terjadi dinegara - negara yang sedang berkembang.

Oleh karena itu sudah dilakukan beberapa penelitian klasifikasi dan prediksi tentang gangguan kelahiran dan gangguan pada ibu hamil menggunakan berbagai algoritma klasifikasi 1,3,5,6,7. Penelitian yang dilakukan Ari Puspita dan Mochamad Wahyudi melakukan prediksi kelahiran prematur pada bayi menggunakan Algoritma C4.5 berbasis Decision Tree dengan menggunakan beberapa atribut atau fitur data sebagai penentu hasilnya [1]. Penelitian Finki Dona Marleny mengenai faktor yang mempengaruhi penyakit asfiksia pada bayi yang baru lahir menggunakan Multilayer Perceptron Neural Network [3]. Penelitian Ari Puspita melakukan prediksi kelahiran bayi prematur dengan Algoritma C4.5 berbasis Particle Swarm Optimization dengan menambahkan Particle Swarm Optimization untuk optimasi pada Algoritma C4.5 [5]. Penelitian Retno Arie Agustien melakukan pemodelan resiko berat badan bayi lahir rendah berdasarkan pendekatan multivariate adaptive regression spline [6]. Penelitian Mambang dan Agus Byna melakukan analisa perbandingan algoritma C4.5, Random Forest dengan Chaid decision Tree untuk klasifikasi tingkat kecemasan ibu hamil [7].

Seleksi merupakan salah satu cara untuk mengurangi beberapa fitur dalam proses klasifikasi dengan tujuan untuk mengurangi dimensi dari data yang kita gunakan dalam proses penelitian. Proses ini juga dapat mengurangi beberapa fitur yang kurang relevan dalam data dengan tujuan untuk meningkatkan akurasi dari proses klasifikasi. Information gain merupakan salah satu algoritma seleksi fitur yang banyak digunakan dan populer. Salah satu kelebihan dari information gain adalah baik digunakan dalam memilih atribut khususnya dalam menangani data dengan dimensi data. Proses klasifikasi dengan atribut yang terlalu banyak jelas akan memerlukan biaya komputasi yang mahal. Terlebih lagi jika beberapa atribut yang digunakan tidak relevan yang dapat membuat performa algoritma klasifikasi menurun.

Dengan masih banyaknya bayi yang lahir secara prematur pada tiap tahunnya, maka pada penelitian ini dibuatlah sistem untuk klasifikasi kelahiran prematur dengan

algoritma C4.5 dan untuk mendapatkan hasil yang terbaik dengan akurasi yang baik maka akan dilakukan analisa terhadap beberapa *Seleksi Fitur* yang akan digunakan. Dengan menggunakan *Seleksi Fitur* yang memiliki akurasi terbaik dalam menyeleksi fitur, yang dilihat dari penelitian sebelumnya tentang stabilitas *seleksi fitur* yang terbaik sehingga akhirnya nanti menghasilkan klasifikasi yang memiliki *accuracy* yang baik.

1.2 Rumusan Masalah

Dari uraian latar belakang diatas, dapat dirumuskan masalah sebagai berikut :

- a. Bagaimana cara melakukan penyeleksian fitur yang digunakan untuk proses klasifikasi.
- b. Bagaimana cara mengimplementasikan Teknik Information Gain dan Correlation – based Feature Selection (CFS).
- c. Bagaimana melakukan klasifikasi dengan menggunakan Algoritma C4.5
- d. Bagaimana mengetahui perbandingan *accuracy* dari hasil klasifikasi C4.5 dan teknik Information Gain dan Correlation – based Feature Selection (CFS)

1.3 Batasan Masalah

Agar pembahasan masalah ini dapat terarah, maka perlu adanya batasan masalah. Maka perlu ditetapkan batasan – batasan masalah yang akan dibahas, yaitu :

- a. Fitur yang akan digunakan untuk diseleksi berjumlah 10 fitur dan data yang akan digunakan berjumlah 500 data yang mana nantinya akan dibagi menjadi 70% Data Train dan 30% Data Test.
- b. Teknik *Seleksi Fitur* yang digunakan dalam penelitian ini adalah metode Information Gain dan Correlation – based Feature Selection (CFS).
- c. Menggunakan algoritma C4.5 untuk proses klasifikasi setelah *seleksi fitur* menggunakan Information Gain dan Correlation – based Feature Selection (CFS).
- d. Data yang digunakan adalah dataset penyakit gynekologi dalam format excel.

1.4 Tujuan

Berdasarkan perumusan masalah, maka tujuan penelitian ini adalah :

- a. Melakukan penyeleksian fitur dengan menggunakan metode Information Gain dan Correlation – based Feature Selection (CFS).
- b. Melakukan Teknik *Seleksi Fitur* dengan metode Information Gain dan Correlation – based Feature Selection (CFS).
- c. Melakukan klasifikasi dengan algoritma C4.5 dengan Fitur yang telah diseleksi.

- d. Menghitung *accuracy* dari *Seleksi Fitur* dan Algoritma C4.5 dalam klasifikasi.

1.5 Metodologi Penelitian

Dalam mengerjakan penelitian ini menggunakan algoritma C4.5 dengan menerapkan *seleksi fitur Correlation – based feature selection dan Information Gain*. Algoritma C4.5 merupakan algoritma yang membentuk suatu pohon keputusan untuk penyelesaian masalahnya. Sedangkan tujuan penerapn *seleksi fitur* pada algoritma C4.5 adalah untuk meningkatkan hasil dari klasifikasi. Adapun 7 tahapan kerja yang dilakukan penulis adalah sebagai berikut :

1.5.1 Studi Pustaka

Studi pustaka merupakan tahapan dimana peneliti untuk memahami topik yang ingin diteliti, dengan cara menghipun informarsi yang relevan, melalui buku – buku ilmiah, laporan penelitian, karangan – karangan ilmiah, tesis dan disertasi, peraturan – peraturan, ketetapan – ketetapan, buku tahunan, ensiklopedia, dan sumber – sumber tertulis baik tercetak maupun elektronik. Pada tugas akhir ini, studi pustaka yang dilakukan meliputi pemahaman mengenai penggunaan metode *Seleksi Fitur* dan Klasifikasi menggunaka Algoritma C4.5.

1.5.2 Pengumpulan Data

Tahap selanjutnya adalah melakukan proses pengumpulan data berupa data set penyakit gynekologi dalam bentuk format excel yang didapatkan dari penelitian sebelumnya yang bersumber dari Rumah Sakit RSUPN Cipto Mangunkusumo Jakarta.

1.5.3 Analisa Sistem

Pada tahap ini dilakukan penelitian bertujuan untuk mengetahui tingkat pengaruh fitur – fitur yang digunakan untuk prediksi kelahiran prematur. Kemudian menggunakan metode seleksi fitur dan Algoritma apa saja yang akan digunakan pada penelitian ini.

1.5.4 Desain Sistem

Dalam tahapan ini penulis melakukan desain secara lengkap sesuai dengan tahapan analisa yang sebelumnya. Dalam desai sistem ini juga menggambarkan alur dari proses klasifikasi menggunakan algoritma C4.5 dan proses *seleksi fitur* menggunakan *Correlation – based feature selection dan Information Gain*.

1.5.5 Implementasi

Tahap implementasi berupa pengolahan data dari dataset ke Data Train dan Data Test berdasarkan penjelasan – penjelasan pada bagian – bagian sebelumnya beserta hasil pengujiannya. Implementasi dilakukan berdasarkan hasil analisa dan desain sistem.

1.5.6 Pengujian Sistem

Tahap ini dilakukan pengujian sistem dimana sistem melakukan teknik Seleksi Fitur dengan menggunakan Data Test. Dari hasil *Seleksi Fitur* tersebut dilakukan evaluasi sistem dengan menghitung *accuracy* dari Algoritma C4.5 pada klasifikasi berdasarkan metode Correlation based Feature Selection dan Information Gain yang digunakan. Selanjutnya menarik kesimpulan dari hasil penelitian yang telah dilakukan.

1.5.7 Pembuat Laporan

Ketika tahap – tahap sebelumnya sudah selesai dikerjakan, dimana tahapan ini bertujuan untuk pembuatan laporan dan dapat dipergunakan untuk pengembangan atau penelitian lanjutan.

1.6 Sistematika Penulisan

Sistematika penulisan skripsi ini secara garis besar akan diuraikan sebagai berikut :

BAB 1 : Pendahuluan

Bab ini menjelaskan mengenai latar belakang, rumusan masalah, batasan masalah, tujuan, metode penelitian dan sistematika penulisan.

BAB 2 : Landasan Teori

Bab ini menjelaskan tentang teori – teori yang menjadi acuan dalam penulisan skripsi yaitu mengenai tahap dalam perancangan basis data yang diperoleh dari beberapa buku literatur, perpustakaan dan internet.

BAB 3 : Perancangan Sistem

Bab ini menguraikan tentang analisa sistem berjalan dan perancangan sistem yang akan diterapkan.

BAB 4 : Implementasi dan Pengujian Sistem

Bab ini berisi tentang perangkat implementasi program dan pengujian program dari sistem penerjemah yang dibangun.

BAB 5 : Kesimpulan dan Saran

Bab ini berisi kesimpulan dari hasil analisa bab–bab sebelumnya. Serta memberikan saran–saran untuk mengembangkan suatu aplikasi yang lebih efektif.

